A temporal landmark for syllabic representation of continuous speech in human superior temporal gyrus

Yulia Oganian, Edward F. Chang

Understanding continuous speech requires an efficient mechanism for segmentation of the speech signal into single syllables (e.g. /seg/-/men/-/ta/-/tion/). A widely held theory is that speech syllabification relies on slow modulations of the amplitude envelope of speech, as speech amplitude peaks in syllabic centers (vowels) and reaches local minima around syllabic boundaries. It was thus suggested that neural segmentation of speech into syllables is based on detection of troughs or peaks in the envelope. In contrast, animal auditory neurophysiology suggest that neural responses might selectively represent auditory edges, i.e. rapid increases in the envelope, or the continuous moment-by-moment envelope itself. It is currently unclear, and intensively debated, how the speech envelope is represented in human non-primary auditory cortex on the superior temporal gyrus (STG), limiting the advancement of speech perception theories.

Here, we combined high-density intracranial recordings, computational modelling, linguistic and acoustic analysis to determine how speech is represented in human STG, and how this representation gives rise to the detection of syllables.
In Experiment 1, participants (n = 11) passively listened to a wide range of natural sentences. Responses were best characterized as a series of evoked responses to discrete events in the envelope, not encoding of the continuous moment-by-moment envelope. However, due to the rapid dynamics of natural speech, it was not possible to determine what discrete events drove the neural response.

In Experiment 2, a subset of participants (n = 3) listened to artificially slowed sentences. In slowed speech, the time intervals between envelope peaks, troughs, and auditory edges were sufficiently large such that neural responses could be clearly attributed to a single event. Neural responses to speech amplitude modulations were best characterized as evoked responses to onset edges in the speech signal, defined as local peaks in the first derivative of the speech envelope (peakRate). In contrast, we found no evidence for encoding of troughs or peaks in the envelope.

To confirm that neural responses reflected the temporal dynamics of amplitude rises at auditory edges, participants (n = 8) in Experiment 3 listened to amplitude modulated non-speech tone stimuli. Neural responses were time-locked to onsets of amplitude ramps and were larger for faster amplitude rises. This provides direct evidence for the preferential encoding of relative change in amplitude, time-locked to local peaks in the rate of change of the amplitude.

Finally, acoustic analysis of natural speech revealed that auditory onset edges reliably cue the information-rich transition between the onset (consonants) and nucleus (vowels) of syllables. Furthermore, we found that the magnitude of peakRate – encoded

in the magnitude of neural responses - cued whether or not a syllable was lexically stressed. Collectively, our findings establish that encoding of auditory edges in superior temporal gyrus underlies the perception of the syllabic structure of speech.

Tracking stimulus statistics from sensory cortices to frontal cortex

Jennifer Lawlor , Celian Bimbard, Shihab Shamma, Yves Boubenec

Complex, cluttered acoustic environment, such as a busy street, are characterised by their ever-changing dynamics. Despite their complexity, listeners can readily detect changes in continuous acoustic streams. However, the neural basis of the extraction of relevant information in complex continuous streams for goal-directed behavior is currently not well understood. As a model for change detection in complex auditory environments, we designed spectrotemporally broad tone clouds whose statistics change at a random time. Ferrets were trained to detect these changes. Hence, they are faced with the dual-task of estimating the baseline statistics and detecting a potential change in those statistics at any moment, mimicking real-life challenges. To characterize the extraction and encoding of relevant sensory information along the cortical hierarchy, we performed electrophysiological recordings in the primary auditory cortex (A1), secondary auditory cortex (PEG) and frontal cortex (FC) of the behaving ferret. A1 neurons exhibited strong onset responses and change-related discharges specific to neuronal tuning. PEG population showed reduced onset-related responses, but more categorical change-related modulations. Finally, a subset of FC neurons (dlPFC/premotor) presented a generalized response to all change-related events only during behavior. We show using a GLM the same subpopulation in FC encodes for the sensory and decision signal. In addition, PEG and FC population dynamics showed a time-dependent evolution within trials and before the change occurred, possibly reflecting an online estimation of the ongoing baseline statistics, a necessary process in this dual-estimation task. All together, these area-specific responses suggests a behavior-dependent mechanism of sensory extraction and generalization of task-relevant event.

Illusory sound texture reveals statistical completion in auditory scene analysis

Richard McWalter, Josh H. McDermott

Auditory scenes often contain multiple sound sources that vary in their temporal homogeneity. Sound textures, as arise from falling rain, galloping horses or swarming insects, lie on one end of this spectrum, having properties that remain relatively constant over modest time windows. Sound textures are well described by statistics measured from early auditory representations [McDermott & Simoncelli, 2011], and unlike speech and other non-stationary sounds are thought to be represented with statistics averaged over a multi-second window. However, little is known about their perception when present in auditory scenes containing other sources.

We investigated the perception of textures when they co-occur with other sounds that could intermittently mask the texture, making it inaudible for segments of time. We asked whether the auditory system infers the presence of background texture when it might be masked by such other sounds. We observed that when textures were interrupted by several seconds of noise, they were heard to continue during the noise provided it was sufficiently high in intensity. The effect is analogous to the well-known "continuity illusion" that occurs for tones interrupted by noise, but differs in lasting much longer than the effect for tones (~2 seconds vs. ~200 ms), and because the extrapolated sound must be defined statistically due to the stochastic nature of texture.

We next asked whether the representation of illusory texture is similar to that of actual texture. In a second experiment listeners judged which of two textures was most similar to a reference texture [McWalter & McDermott, 2018]. The first texture underwent a change in statistics during its duration, while the second texture had fixed statistics. The experiment leveraged the fact that estimates of texture statistics are biased by several seconds of the stimulus history. We found that when interrupting noise was inserted prior to the statistic change, texture judgments were biased in a manner comparable to that if texture was actually present, suggesting that the illusory texture heard during the noise is represented like actual texture. When a silent gap was introduced prior to the noise, which disrupted the subjective impression of illusory texture, the bias was eliminated.

The results suggest that illusory sound textures can be heard over several seconds and appear to be represented similarly to actual texture, revealing new aspects of perceptual completion in auditory scenes.

Amygdala-TRN projections amplify tone-evoked activity in auditory thalamus and cortex

Solymar Rolon-Martinez, Mark Aizenberg, Maria N. Geffen

Many forms of behavior require selective amplification of neuronal representations to relevant sensory signals. Associating emotional responses with sensory cues can lead the nervous system to alter behavior to future representations of these cues. Here, we identify a novel pathway between the baso-lateral amygdala (BLA), an emotional learning center in the mouse brain, and the inhibitory nucleus of the thalamus (TRN), and demonstrate that activation of this pathway amplifies sound-evoked activity in the central auditory pathway. We stimulated BLA using channelrhodopsin (ChR2) with a laser via implanted optic cannulas, while recording neuronal activity in the auditory cortex (AC) in response to a presentation of random tone sequences in awake, head-fixed male or female mice. Optogenetic activation of the BLA suppressed spontaneous activity (paired t-test, p=0.0007), while amplifying tone-evoked response magnitude in AC (paired t-test, p=8.5e-5, n= 8 mice). Inspection of fluorescence following RetroBead injections in TRN revealed direct projections from BLA to TRN. These projections were further confirmed by retrograde labeling of neurons in the BLA using a CAV-2 virus in TRN. We next directly activated projections from the BLA to TRN by repeating the initial experiment, but positioning the optic cannula over TRN. We found that there was a significant suppression of spontaneous activity (paired t-test, p=0.003, n= 7 mice), and a significant increase in tone-evoked responses in AC (paired t-test, p=3.9e-8). We found that activation of the BLA projections to TRN also led to inhibition of spontaneous activity (paired t-test, p=4.3e-9) and an increase in tone-evoked responses in auditory thalamus (Medial Geniculate Body, MGB) (paired t-test, p=3.4e-7, n = 5 mice), consistent with the hypothesis that the changes in AC responses with BLA activation are a result of projections from BLA to TRN via MGB. These results demonstrate a novel circuit mechanism for amplification of sensory representation of behaviorally relevant signals and provide a potential target for treatment of neuropsychological disorders, in which emotional control of sensory processing is disrupted.

Dissociation of task engagement and arousal effects in auditory midbrain and cortex

Daniela Saderi, Charles R. Heller, Zachary P. Schwartz, Stephen V. David

It is well-established that changing behavior state has diverse effects on auditory brain activity, but these findings have yet to be integrated into a coherent theory of how internal state influences sensory coding. Many previous studies have focused on the effects of a single state variable, and interpretation of these results is complicated if the variable of interest is correlated with other, uncontrolled changes in state. To isolate effects of different state variables on auditory processing, we developed a paradigm to simultaneously control task engagement and monitor fluctuations in arousal during single-unit recording from the inferior colliculus (IC) and primary auditory cortex (A1) of ferrets. To control task engagement, neural activity was recorded during a tone versus noise discrimination task and during passive presentation of task stimuli. Arousal was measured via pupillometry.

We used a generalized linear model to isolate the effects of engagement and arousal variables on spontaneous and evoked activity in the IC and A1. As expected from previous studies, fluctuations in pupil-indexed arousal were correlated with changes in task engagement, but their effects could be dissociated in most data sets. In both areas, individual neurons could be modulated by either engagement or arousal or by both variables. However, engagement and arousal effects in the IC were about half the magnitude of those in A1. Engagement effects could be enhancing or suppressing in both areas. Arousal effects also had variable sign in IC, but arousal was mostly positively correlated with spike rate in A1. Taking both state variables into account revealed a smaller measured influence of task engagement than when arousal was not considered. These results indicate that some changes attributed to task engagement in previous studies should in fact be attributed to global changes in arousal. Moreover, these arousal effects may explain differences in neural activity observed between passive conditions pre- and post-behavior. This same approach can be used to account for other state variables, such as selective attention and behavioral effort, providing a general method for dissociating the influence of continuous and discrete behavioral state variables.

Amodal neural representations

Muers RS, Perrodin C, Kayser C, Chandrasekaran C, Kocsis Z, Kikuchi Y, Wilson B & Petkov CI

Amodal representations are those associated with unique entities that are consistent across any form of sensory input (e.g., hearing the voice of an individual, seeing their face or both). Such representations are important for forming abstract concepts. A dominant hypothesis advanced from human neuroimaging and neuropsychological data is the "hub-and-spoke" model, whereby modality-specific information from sensory cortices ("spokes") feeds into modality invariant convergence sites ("hubs"). However, the neurophysiological bases for amodal representations are unknown, and we examined the hypothesis that neural amodal representations are distributed across auditory cortex and sensory convergence sites. We studied neural responses in the macaque monkey auditory cortex on the supratemporal plane (STP) and the upper bank of the superior temporal sulcus (STS). The monkeys were presented with auditory, visual, and audio-visual stimuli, including voices, faces, and voice-face combinations and other naturalistic stimuli. Amodal responses were identified using a vector similarity analysis of local-field potential (LFP) or spiking firing rate responses to the stimuli in any sensory modality. The Euclidean distance between all combinations of auditory (A), visual (V) and audio-visual (AV) stimuli (|A-V|, |A-AV|, |V-AV|) was cast as a vector and converted to a normalised Amodal Similarity Index (ASI). A response component was categorised as amodal if the ASI value was <5%, indicating a high level of similarity across all stimulus combinations for a specific individual. Results from neural activity quantified by LFPs in time and frequency showed a substantial proportion of recording sites in both the STP and STS that displayed amodal responses (37.4% and 31.8%, respectively). However, the time-frequency patterns were strikingly different in the two regions: auditory STP sites showed amodal response components associated with gamma activity (35-70Hz) throughout and after stimulus presentation, whereas amodal responses in the STS were associated with earlier lower frequency components (<15Hz). In conclusion, amodal representations appear to be more broadly distributed than suggested by the "hub-and-spoke" model and are differently reflected in neurophysiological response patterns in auditory and association cortices.